Springboard DSC Program

Capstone 2 Milestone report

Screening for Dyslexia Using Eye Tracking During Reading

Using Eye Movement during Reading to predict Dyslexia

By Harinee Madhusudhan

June 2020

Although dyslexia is fundamentally a language-based learning disability, our results suggest that eye movements in reading can be highly predictive of individual reading ability and that eye tracking can be an efficient means to identify children at risk of long-term reading difficulties.

Our study is based on a sample of 97 high-risk subjects with early identified word decoding difficulties and a control group of 88 low-risk subjects. These subjects were selected from a larger population of 2165 school children attending second grade.

The objective was to create simple and commonly used machine learning models like logistic regression, and maybe others to compare how they perform regarding for the task of predicting the high risk of long- term reading disabilities.

Description of the dataset, how obtained, cleaned, and wrangled it

The data is available on FigShare, here:

https://figshare.com/collections/Screening_for_Dyslexia_Using_Eye_Tracking_During_Reading/3521379/1

By tracking eye movements during reading, the observant was able to follow the reading process as it occurs in real-time and obtain objective measurements of this process. The data being sampled provide a next to continuous record of reading that reflects both the speed and accuracy of the processes involved. Importantly, this mode of measurement requires no overt response extraneous to the reading process itself and thus makes it possible to assess reading performance without placing additional task demands on the subject.

	DistanceL	DistanceR	Gender	Label	Subject	LTypeFSum	LTypeSSum	LTypeFCount	LTypeSCount	LTypeFMean	...	LVerDirNMean	RVerDirUSu
0	13413.296257	13580.306540	1	0	111GM3	15320.0	14520.0	171	172	89.590643	...	62.027027	1122
1	8788.682167	8509.062191	0	1	111JA2	22100.0	17840.0	288	288	77.003484	...	43.737024	1374
2	9765.357380	10281.893102	1	1	111RP1	24780.0	15180.0	239	240	103.682008	...	47.918367	1468
3	11950.957324	11461.339153	1	0	112JU3	23920.0	15960.0	243	244	98.436214	...	46.666667	1060
4	4959.743932	4913.022136	1	1	112KA1	29920.0	9760.0	249	249	120.645161	...	33.717579	1242

The distance is computed using sqrt(square(x1-x2) + square(y1-y2). From the data extract few more columns from the eye movement data by adding LType and RType. S-Saccade and F-Fixation: If the distance travelled is < 1 (review this) then it is Fixation else Saccade. Also, the eye movements of left and right in the direction of horizontal and vertical. Here three displacements are considered one left, right and no change (L, R, N) and (U, D, N). Also, Since the data is recorded per folder, we used a loop function to reach out to each folder and collect them all in to one table. That is a data frame with 185 rows x 53 columns

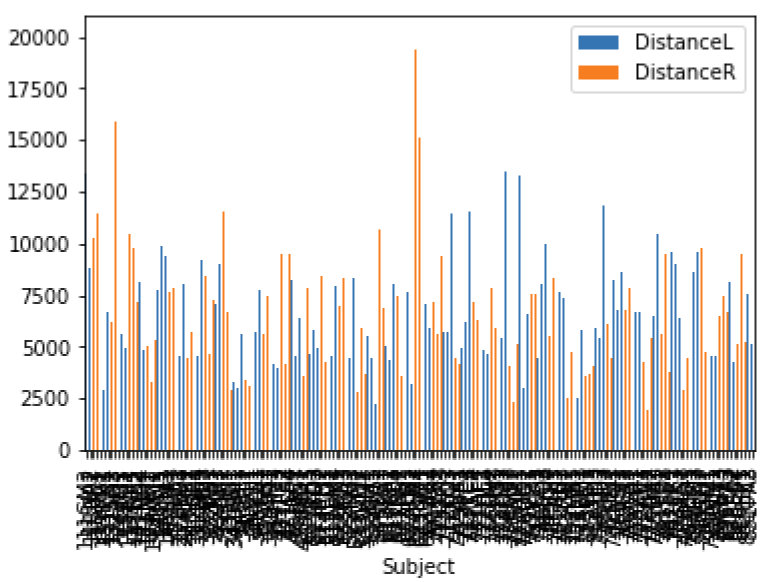
```
Index(['DistanceL', 'DistanceR', 'Gender', 'Label', 'Subject', 'LTypeFSum',
       'LTypeSSum', 'LTypeFCount', 'LTypeSCount', 'LTypeFMean', 'LTypeSMean',
       'RTypeFSum', 'RTypeSSum', 'RTypeFCount', 'RTypeSCount', 'RTypeFMean',
       'RTypeSMean', 'LHorDirLSum', 'LHorDirRSum', 'LHorDirNSum',
       'LHorDirLCount', 'LHorDirRCount', 'LHorDirNCount', 'LHorDirLMean',
       'LHorDirRMean', 'LHorDirNMean', 'RHorDirLSum', 'RHorDirRSum',
       'RHorDirNSum', 'RHorDirLCount', 'RHorDirRCount', 'RHorDirNCount',
       'RHorDirLMean', 'RHorDirRMean', 'RHorDirNMean', 'LVerDirUSum',
       'LVerDirDSum', 'LVerDirNSum', 'LVerDirUCount', 'LVerDirDCount',
       'LVerDirNCount', 'LVerDirUMean', 'LVerDirDMean', 'LVerDirNMean',
       'RVerDirUSum', 'RVerDirDSum', 'RVerDirNSum', 'RVerDirUCount',
       'RVerDirDCount', 'RVerDirNCount', 'RVerDirUMean', 'RVerDirDMean',
       'RVerDirNMean'],
      dtype='object')
```

Exploratory Visualization

While the subjects were attending 3rd grade (age 9–10), eye movements were recorded as part of an ophthalmological examination that aimed to investigate whether there were any differences between the two groups in terms of basic visual and oculomotor functions. The eye movement recordings were made while the subjects were reading a short natural passage of text adapted to their age. Recordings were available for 185 subjects, 97 HR(High-risk) subjects (76 males and 21 females) and 88 LR(Low risk) subjects (69 males and 19 females). All subjects read one and the same text presented on a single page of white paper with high contrast. The text was distributed over 8 lines and consisted of 10 sentences with an average length of 4.6 words.

```
df.plot(kind='bar', y=['DistanceL', 'DistanceR'], x='Subject')
```

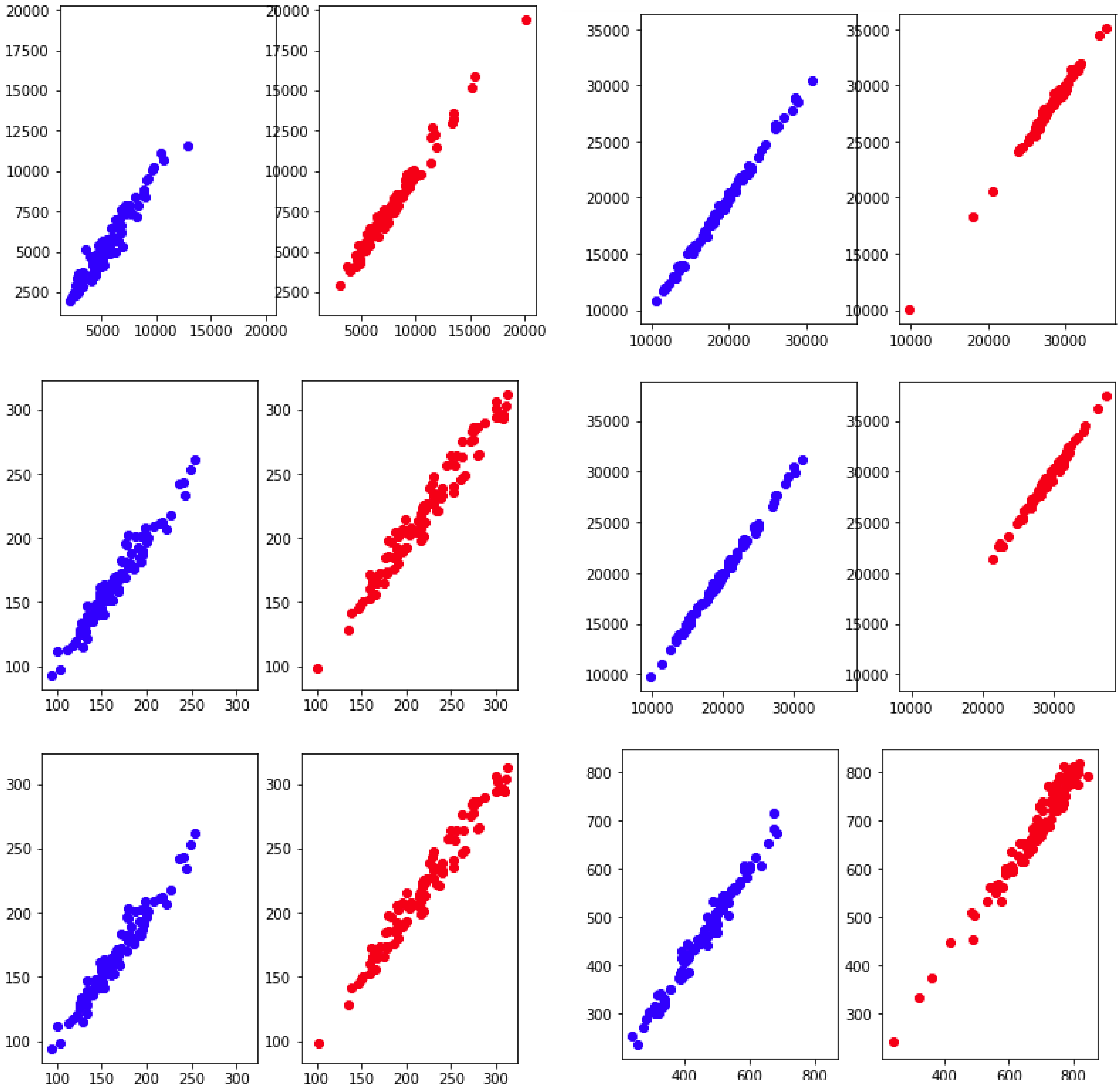
```
: <matplotlib.axes._subplots.AxesSubplot at 0x2072f589898>
```



Eye Movement Analysis

The dataset consists of numerical values from the transformed features which has Gender, Label, Subject and 50 features derived from the movement of the eye and time that were recorded. Since it is derived from the data set there is no null or missing values. The information contained in these features captures different quantitative properties of eye movements in reading, including their duration, amplitude, direction, stability, and symmetry.

Comparing left eye to right eye for controls (blue) and dyslexic (red) - No of Fixations, No of Saccades, and distance travel. And Comparing left eye to right eye for controls (blue) and dyslexic (red) - horizontal movement and vertical movements.



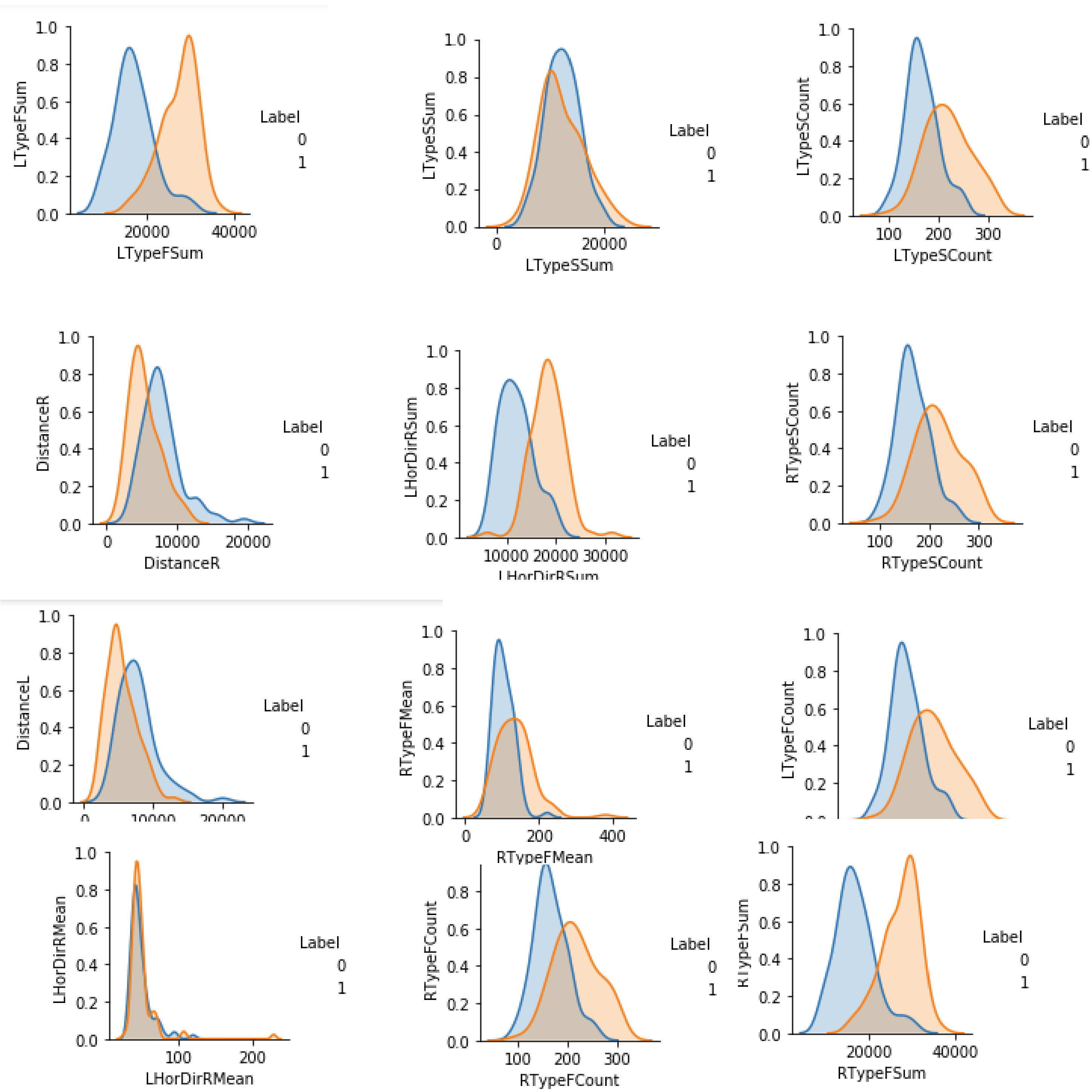
Looking at a scatter plot between the left and right eye movements - distance travelled, we see that both eyes travel almost equal distances, with some exceptions for dyslexic kids, their eyes traveling longer distances.

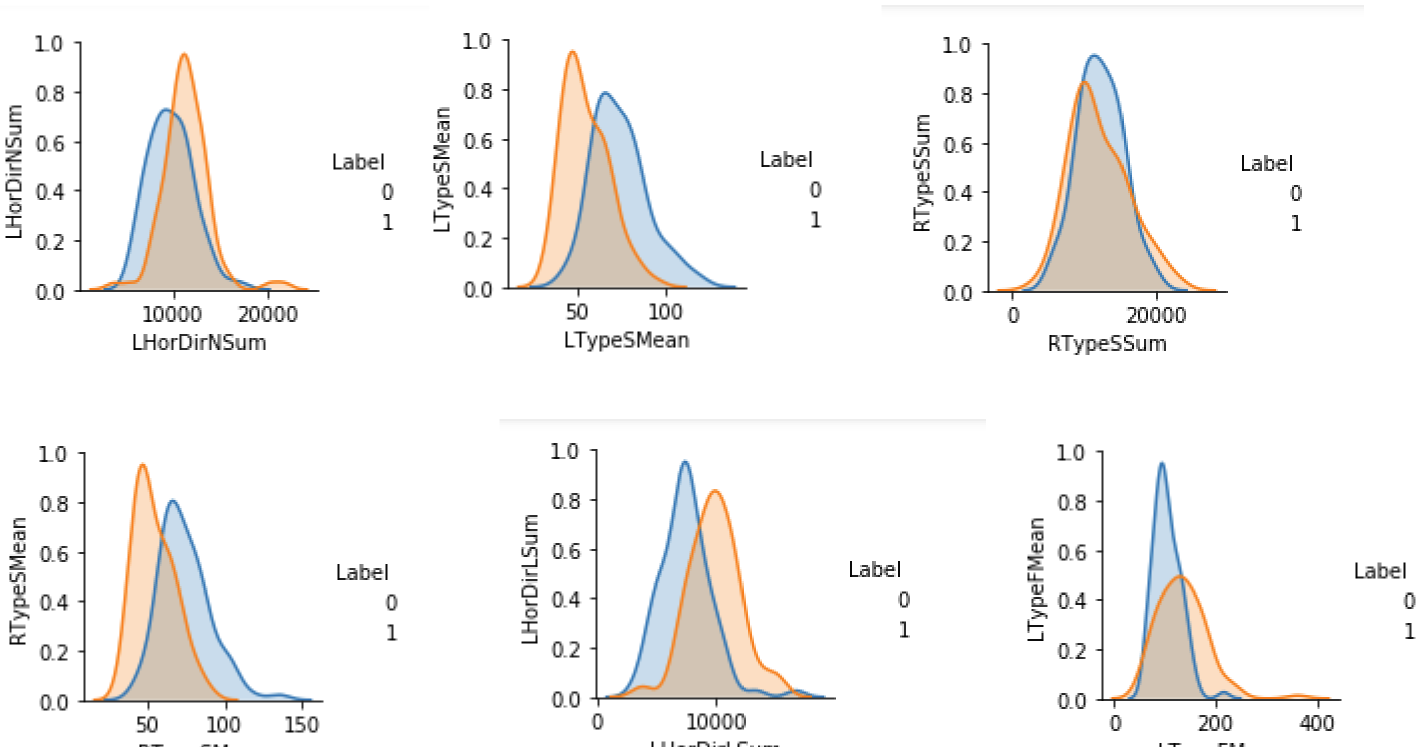
Then we look at the number of saccades and fixations for each of the groups. we see that the number of saccades as well as the fixations are almost the same between the right and left eyes - indicated by the slope of the line. However, the number of saccades and fixations are more for dyslexic kids than control kids.

This may be because dyslexic kids move their eyes more frequently than control groups to travel the same distance.

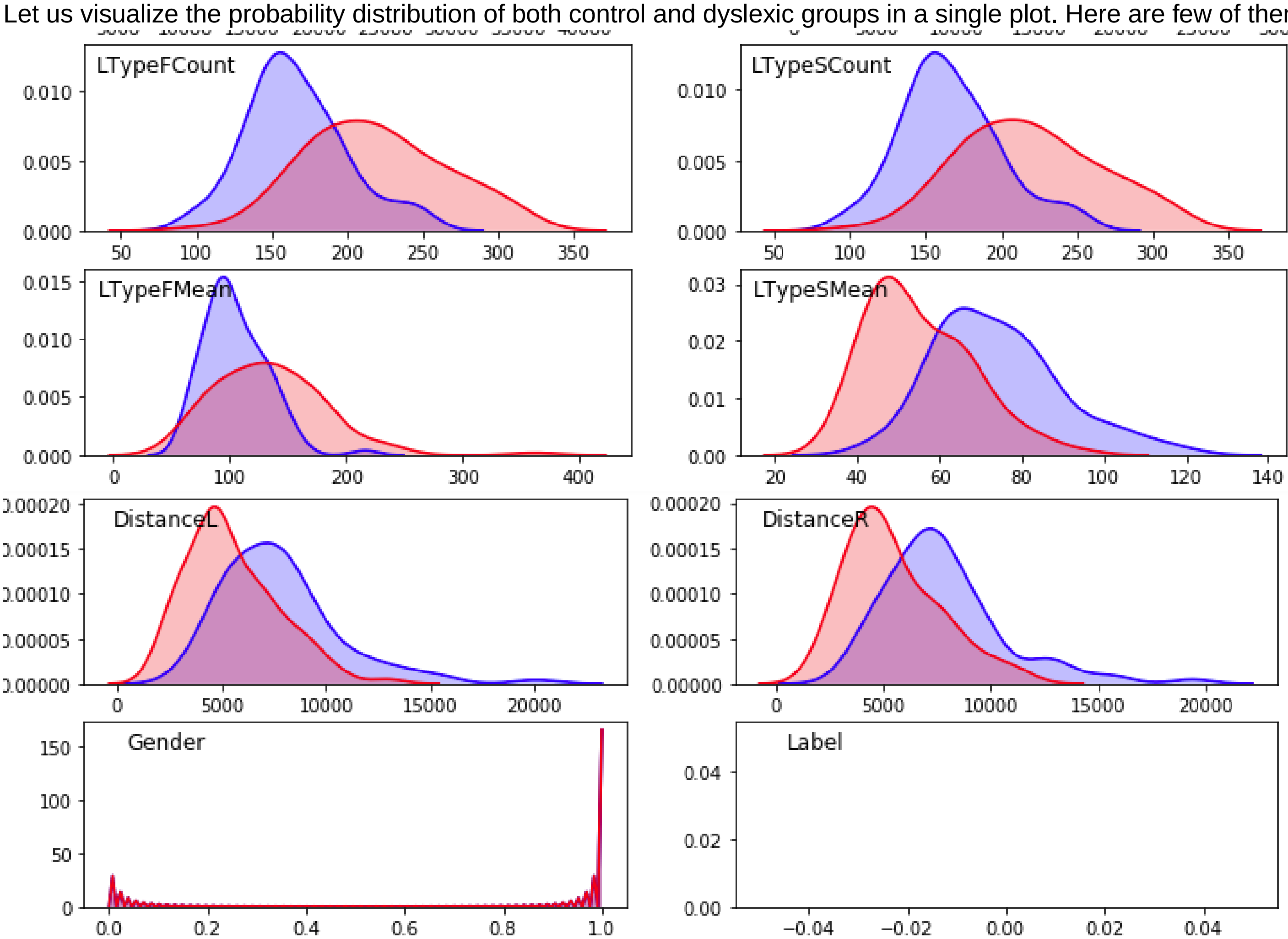
Note: these charts have the same scale for x and y axes

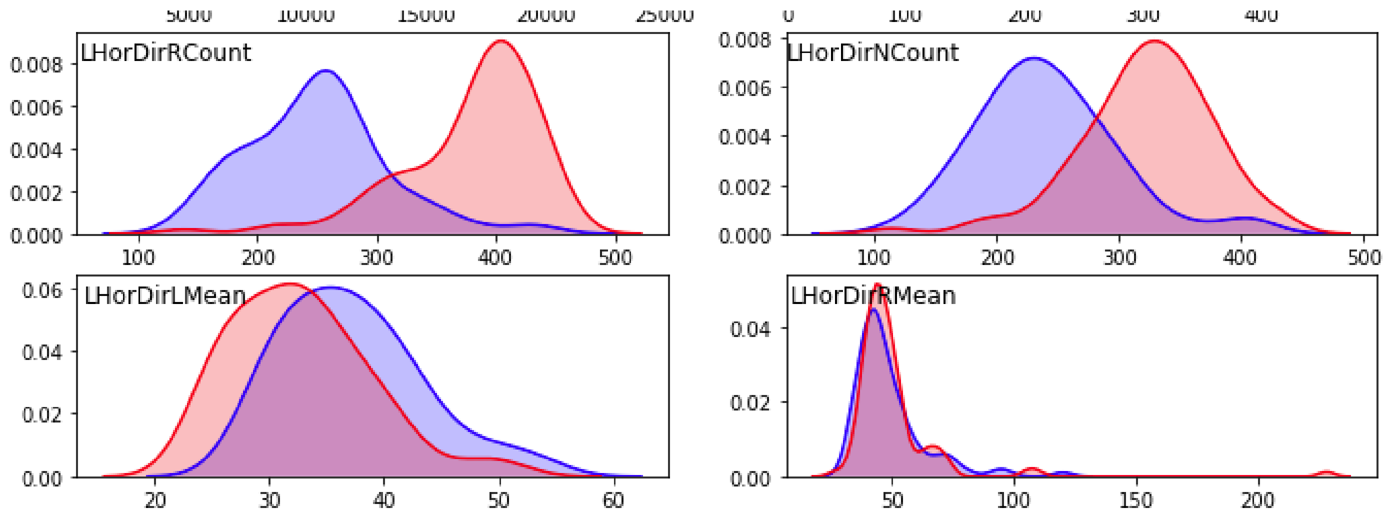
Now Let us compare feature to features for all features.





Let us visualize the probability distribution of both control and dyslexic groups in a single plot. Here are few of them .





Future proposal:

Even though it has been a known fact that the eye movements of dyslexic readers are different from those of typical readers, usually research has focused almost exclusively on identifying group-level differences. The objective here is to use machine learning and predictive modeling, to individual-level predictions with high sensitivity and specificity. For that the different eye movement features are compared and the critical features that differentiate high-risk and low-risk are identified. In short, these models can be used for identifying school children at risk of dyslexia.

I plan to use this data to train two types of binary classification models – a logistic regression, a Gradient Boosted model and a Random Forest model. I plan to use 80% of training and 20% of the data to test. I will also try with a set of 70%-30% training-test to compare the models.

Based on the ROC curve, precision and recall achieved, I will identify one model that performs best to predict dyslexic tendencies based on the eye movements.